Visiphone

Judith Donath **Karrie Karahalios** **Fernanda Viégas**
MIT Media Lab
20 Ames Street
Cambridge, MA 02138
+1 617 253 5098
[judith | kkarahal | fviegas@media.mit.edu]

ABSTRACT
Visiphone is a graphical interface for mediated audio conversations that is designed to support continuous, ubiquitous connections between people in different locations. The graphics show the existence of the audio connection, provide feedback that one's voice is loud enough to carry across the channel, and indicate that someone on the other end of the connection has spoken. They also serve more subtle purposes, providing a focus for attention and visually representing the rhythm of the conversation itself. Our goal is to create an aesthetic object that enables users to perceive conversational patterns that are present but not obvious in traditional communication interfaces.

Keywords
Media spaces, graphical conversation, domestic technology

INTRODUCTION
The scenario: a kitchen. A woman prepares a meal while a friend sits reading at the table. They chat occasionally – the one reading sometimes quotes a particularly interesting passage or the cook mentions some recent event which they discuss for a while and then resume concentrating on their respective tasks. Such casual, companionable conversations are an important part of our domestic life, not only for the information we gather during the course of such an exchange, but also as a means of strengthening social ties – and for the enjoyment we gain from such sociability [3].

Yet today, it is likely that one's close friends and family live quite a distance away. Our goal with this project is to design a communication device that connects two domestic spaces so that people living far apart can still enjoy a companionable, casual conversation; the sort of conversation one has when one is keeping someone company rather than focusing entirely on the discussion. Such a device should leave the hands free to do other tasks, it should allow the user to move freely within a limited area and it should allow several people to participate in the conversation at each end.

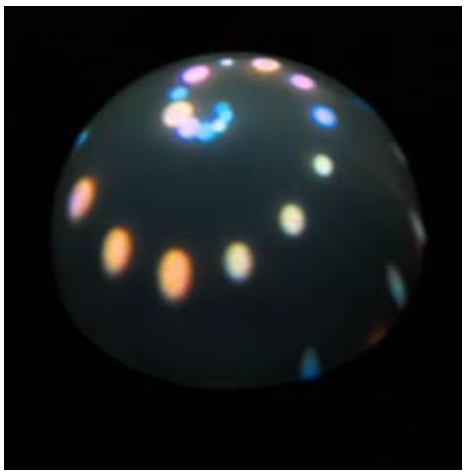

Figure 1. The *Visiphone* display dome.

Our solution is a project we named *Visiphone*, for it is a visual telephone, one that augments speech with graphics. Like a speaker-phone it allows one to talk and listen at some distance from the device, thus allowing hands-free mobile talking and multiple participants. Yet it also provides a visual reminder that the auditory connection exists. And, it provides a dynamic, graphical representation of the rhythm of the conversation which serves as an attention-getting device, gives useful feedback to the speaker, and provides a focal point for directing speech.

PREVIOUS WORK: CONNECTING SPACES WITH AUDIO
There have been a number of "media space" projects that connect geographically distant locales with some combination of audio and video (e.g. [1][9]), as well as studies of the relative affordances of audio, video and other media (e.g. [2][5][8]). Smith and Hudson's work on low disturbance audio found that audio, even when filtered to be incomprehensible (for privacy, in their application) provided a good sense of awareness of the presence and activity of others[9]. The *Somewire* project is one of the most relevant, since it was designed to foster casual interactions among colleagues [8]. Here, Singer et. al experimented with a number of visual interfaces in conjunction with an audio-only media space. They found that control over such features as

localization or other audio attributes was not needed, but that information that supplemented the users' knowledge of the social aspects of the space, such as awareness of the presence of others, was quite welcome and useful.

However, most "media space" work has been done in the context of work environments, which differ significantly from the home in many regards, including privacy requirements, types of on-going activities, and appropriate interface complexity and style. While most studies of technology for the home have tended to focus on labor-saving devices and home automation, some useful ethnographic studies have examined the importance of communication in the domestic sphere and the types of technology that support it [10]; it is becoming increasingly clear that communication is a one of the most valued uses of computer technology in the home.

VISIPHONE

Design goals

Our goal in designing *Visiphone* was to make a device that would enable casual communication between two distant domestic spaces. Unlike an office environment, in which users are generally seated at a desk, people at home move about freely. This is especially true in the kitchen, which is a center of both social activity and household work. We located our design scenarios in this multitasking environment, in which people are engaged in several activities at once and where conversations often ebb and flow as attention is temporarily diverted to other tasks.

One of the key differences between conversations that are conducted face to face and those on the telephone is that the latter tend to be more continuous. In person, a conversation can start and stop, pausing when a topic is exhausted or other tasks beckon; on the telephone, long silences are awkward and uncomfortable – and often signal serious disagreement between the parties. To support more casual conversation, our design would need to help overcome the requirement that the conversation be the unceasingly forefront activity. We identified several possible causes for the telephone conversation's center stage position: cost, the hand-held set, and the phone's audio-only indication of presence.

1. *Cost:* We began with the premise that Internet telephony will soon make it feasible for people to maintain long-distance, long duration audio connections without the anxiety about mounting toll charges that accompany today's long-distance phone calls. *Visiphone* uses the MBone to send audio over the network.

2. *Hand-held set.* The hand-held telephone receiver limits the conversation to a single participant; it also requires the use of one's hand to hold it (or an awkward shoulder shrug). Speaker phones alleviate both of those problems and *Visiphone* works very much like a speaker phone, hands-free and at a distance.

3. *Audio-only indication of presence:* In a phone call, silence is often met with the response "Hello? Hello? Are you still there?". When speech ceases, there is the sense that the connection has been lost. We hypothesize that people feel compelled to maintain continuous conversation on the telephone in part to maintain the perception of a live connection. *Visiphone's* interface is in motion as long as there is a connection; our intent here was to provide a separate, intuitive indicator of connectivity that would relieve the audio channel of this responsibility.

A hands-free, long-term audio connection alone would makes it possible to walk around a room and talk to a distant friend at will. Yet such a system would also have an unfortunate surveillance-like quality: it would be easy to forget that one's home was actually a portal to another space and that all of one's comments were being heard elsewhere. A visible interface that indicates the existence of the live connection alleviates this problem, serving as an ongoing reminder of the audio link.

Another function of the visible interface is to serve as a focus of attention. Anyone who has attended a meeting at which some participants are present via speaker-phone knows the phenomenon by which the little black object on the table becomes a stand-in for the person: one addresses one's remarks to it, one looks at it when it is "speaking". We focus on the object because it is difficult to converse with a disembodied voice, to address one's remarks to the ether. With the traditional speaker-phone, the physical object provides no information, nor is it designed to evoke a person, conversation, connection, etc. With *Visiphone*, we sought to make an object that is abstract enough to be suitable to all users, contexts, and topics, yet that also reflects the rhythm of the ongoing conversation.

Two further considerations influenced our design. The system needed to be easy to use: no complex controls, no custom installation. And, it needed to be a decorative object, one that people would welcome as a featured object in the home.

Description

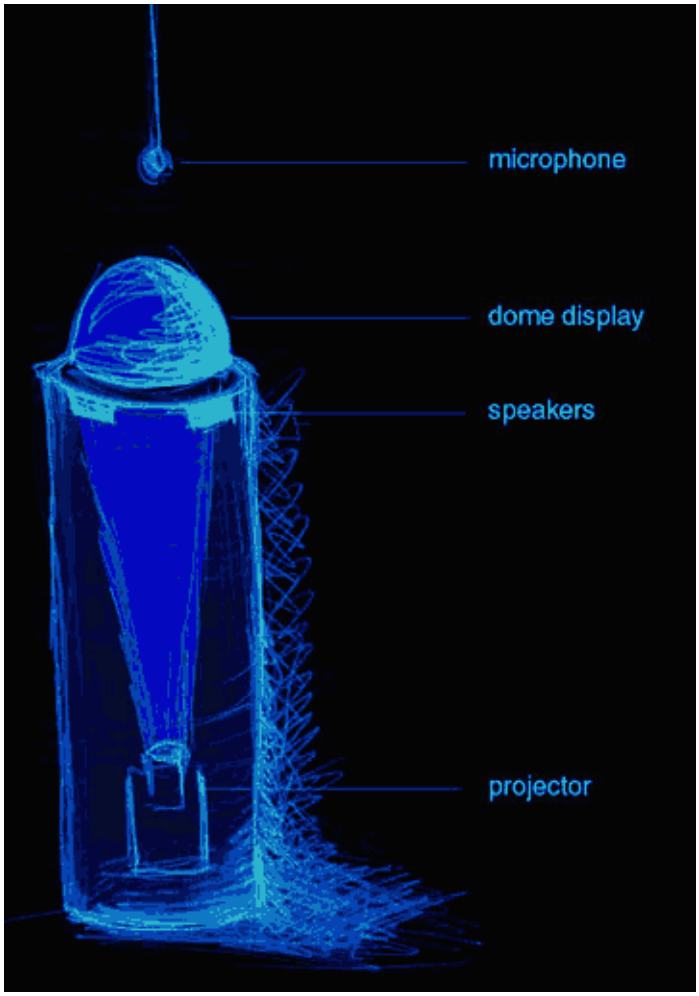

Figure 2. The *Visiphone* installation.

Our final design for the *Visiphone* system consists of two stations connected via the Mbone. Each station has a dome on which the visualization is projected (see *Fig. 2*). When a live connection exists, the dome displays a continuous moving spiral of circles. The central dot represents the present moment. If it is a small gray dot, there is no sound going between the two spaces. When sound is originating locally, the current circle is orange; when sound originates at the outside location the circle is blue. The size of the circle is proportionate to the volume of the audio. If sound is coming from both locations, the colors are shown as concentric, blended circles. The dots spiral outward from the center, so the display shows the history of the last half minute or so of conversational rhythm.

Visiphone's display is a translucent dome sitting on a pedestal (see *Fig. 2)*. The graphics are projected into the dome from below. The dome shape makes it an interface in the round: one can view it from any side. This is essential for an object meant to create a connection between two inhabited, real-world spaces in which people move about. The design of the dome itself is also a key element in this multimodal interface and its size, location, and appearance influence its use and its ability to portray the sense of awareness and continuous connection in the space.

Visiphone does not add any new data that is not already part of the audio channel, it simply re-presents a subset of that data (source and volume, in temporal slices) in another medium. Although from an information theoretic standpoint the visual data is redundant, from a user's standpoint this reification of the audio serves several purposes. Using an audio-only speaker phone to provide a continuous, long-term connection has several drawbacks: in a noisy environment, it is difficult to know whether one's voice has carried or to know to pay attention to new voices emerging from the phone; long periods of silence make it easy to forget the device, which then takes on the unwanted quality of unobtrusive surveillance. By making the audio visible, *Visiphone* turns the speakerphone into a portal between spaces. The existence of the connection is recognizable even during silence and the dynamics and inflections of the conversation are made salient by the abstract visualization. The sudden appearance of vivid dots draws one's attention, even if ambient noise had initially masked the sound. Seeing the color change on the display is a simple but effective mechanism for noting whether one's words can be heard at the other end. *Visiphone*'s graphics do not portray the technical aspects of sound (as do, for example, the audio renderings seen in sound-mixing boards); its purpose is rather to enhance awareness.

DISCUSSION

User response

Although we have not conducted formal user studies of *Visiphone*, it has been on display in public environments[1] and used by hundreds of people. People have been quite enthusiastic about it. The form of the display has proved to be quite important – the spiralling dots are often described as mesmerizing and this aesthetic appeal is an intrinsic part of its value. One of the more surprising comments, but one we heard repeatedly, was that people thought it would make a good "therapy" tool; they were interested in the way the dots could show patterns of interruptions and of individual conversational dominance.

1. It was shown in the Emerging Technologies exhibit at SIGGRAPH '99 and has been a featured demonstration at the MIT Media Lab.

It still remains to test the system in its intended environment, the home. In real use, it would need to compete with the telephone for clarity of sound (our current implementation, using the Mbone for audio, is rather noisy and has some lag). And it remains to be seen in practice how well it accomplishes its stated goal, which is to support casual conversation in an active environment.

Future work

Although we set out to make a very "lightweight" visual interface for audio communication, one that did not require additional information beside the audio signal itself, we have become interested in the question of what information can we add that would help bridge the distance between two domestic environments. Some such data can be extracted from the audio stream alone - for instance, pitch and other features could be used to distinguish among several speakers, giving each their own graphical marker, thus illustrating the conversation in greater detail. Other useful data would need to come from additional sensors. For example, while *Visiphone* shows that a connection is live even during stretches of silence, it does not provide any information about whether anyone is actually in the distal room, or who is there. A "smart room"[7], using vision or other sensors, would be able to provide the system with this information, which could then be visualized.

REFERENCES

1. Bly, S. Harrison, S. and Irwin, S. 1993. Media spaces: Bringing people together in a video, audio and computing environment. *Communications the ACM*, **36**: 1
2. Dourish, P. 1993. Culture and control in a media space. In Proc. Third European Conference on Computer-Supported Cooperative Work ECSCW'93. Milan, Italy.
3. Dunbar, R. 1997. *Gossip, Grooming and the Evolution of Language.* Cambridge, MA: Harvard University Press
4. Hindus, D., Ackerman, M., Mainwaring, S. and Starr, B. 1996. Thunderwire: A field study of an audio-only media space. *Proceedings of CSCS '96.* Boston, MA.
5. Isaacs, E.A. and Tang, J.C. 1994. What video can and cannot do for collaboration: A case study. *Multimedia Systems*, Vol. **2**, 63-73.
6. Mateas, M., Salvador, T. Scholtz, J. and Sorenson, D. 1996. Engineering ethnography in the home. *Proceedings of CHI '96.* Boston, MA.
7. MIT Media Lab Vision and Modeling Group. 1995. Smart Rooms. <http://www-white.media.mit.edu/vismod/demos/smartroom/> Mod: Oct 18, 1995 (21 March 2000).
8. Singer, A., Hindus, S., Stifelman, L., and White, S. 1999. Tangible progress: Less is more in Somewire audio space. *Proceedings of CHI '99.* Pittsburgh, PA.
9. Smith, I. and Hudson, S. 1995. Low disturbance audio for awareness and privacy in media space applications. In *Proceedings of Multimedia '95*. San Francisco, CA.
10. Venkatesh, A. 1996. Computers and other interactive technologies for the home. *Communications of the ACM.* **39**:10.